DOCUMENT RESUME

ED 306 236 TM 012 999

AUTHOR Huntley, Renee M.; Plake, Barbara S.
TITLE An Investigation of Multiple-Response-Option Multiple-Choice Items: Item Performance and Processing Demands.
PUB DATE 88
NOTE 17p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cognitive Processes; Computer Assisted Testing; Difficulty Level; Educational Psychology; Higher Education; *Multiple Choice Tests; Objective Tests; Preservice Teacher Education; *Standardized Tests; Test Construction; *Test Format; Test Items; *Undergraduate Students
IDENTIFIERS American College Testing Program; Item Discrimination (Tests); *Response Patterns

ABSTRACT

The combinational-format item (CFI)--multiple-choice item with combinations of alternatives presented as response choices--was studied to determine whether CFIs were different from regular multiple-choice items in item characteristics or in cognitive processing demands. Three undergraduate Foundations of Education classes (consisting of a total of 120 students) at a large midwestern university participated in this study within their regular classrooms. Items from previous American College Testing programs were modified to form straight linear, single response, or multiple response formats for either single answer or all-inclusive answer items. A second study investigated the processing demands of the combinational-format. Undergraduates (n=43) in educational psychology courses at a large midwestern university were administered a computerized test comparing the formats, and reaction times were measured. Straight linear, single-response, and multiple-response formats showed similar item difficulty and discrimination values regardless of their response option status, indicating no reason to prefer one item phrasing over the others. In the second experiment, combinational options resulted in more processing demands than did non-combinational ones. Although item characteristics did not seem affected by the use of multiple-response items, the CFIs did appear different in the processing demands placed on the subject. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED306236

An Investigation of Multiple-Response-Option Multiple-Choice Items: Item Performance and Processing Demands

Renee M. Huntley

American College Testing Programs

Barbara S. Plake

University of Nebraska-Lincoln

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Renee M. Huntley

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Running head: MULTIPLE-RESPONSE-OPTIONS

TM012999

An Investigation of Multiple-Response-Option Multiple-Choice Items: Item Performance and Processing Demands

Most standardized tests contain items presented in a multiple-choice format. This format generally consists of a stem and four or five alternatives, one of which is considered to be the "correct" or "keyed" option. One variation of this item type is the multiple-choice item with combinations of alternatives presented as response choices. This item type is often found in medical sciences tests for licensure and certification and has been referred to as the "Type K" item by the National Board of Medical Examiners; or the "I, II, III format;" or the "Complex Multiple Choice" item (CMC). This paper will use the term "Combinational-Format Item" to describe this item type. It is believed by some measurement experts that such an item format not only allows for greater flexibility in item development but often creates situations that tap higher-level thought processes.

Research on the Combinational-Format Item has concentrated on comparing this item with multiple true false (MTF) or "Type X" items. For the MTF item, the examinee's task is to make true/false judgments for each of the alternatives that follow the stem. Comparisons of MTF items with Combinational-Format Items almost always show that MTF items are harder and more reliable than their Combinational-Format Item counterparts (Kolstad, Wagner, Kolstad, & Miller, 1983; Albanese, 1982; Harasym, Norris, & Lorscheider,

1980; Albanese, Kent, & Whitney, 1979). However, in each of these comparison studies, the number of potential options to be considered in the Combinational-Format Item version was naturally less than the number of options to be considered by the examinee for the MTF item. For example, in the Combinational-Format Item, the typical number of combinations presented for examinee consideration are four or five, thus fitting into the standard multiple-choice item framework. For the MTF item, if q options are provided, 2**q alternatives must be processed by the examinee. Since the item presents more possible alternatives to be considered, it is not surprising, psychometrically, that such items would be more reliable (Albanese, Kent, & Whitney, 1979).

Most ordinary multiple-choice tests, however, are designed to use four- or five-alternatives; thus the MTF item is not a realistic choice. Given this constraint of four or five multiple-choice alternatives, the appropriate research question is, then, whether the multiple-choice item with four or five combinational-response options is somehow "different" from its four- or five-alternative regular multiple choice item counterpart. It is feasible, as pointed out earlier, that this Combinational-Format Item allows for the measurement of higher-level thought processes. It is also possible that such an item format creates confusion in the students, requires more rereading of the options, or sets up an expectation that a multiple response must be correct (even though, in some

applications, an alternative like "I only" may be presented and possibly be the keyed answer).

The purpose of this study, therefore, is to analyze the Combinational-Format Item in relation to a more compatible multiple-choice item in order to discern whether the Combinational-Format Item is different than the regular multiple-choice item in item characteristics or in cognitive processing demands.

A series of two experiments were run to address these questions. In Experiment I, the same basic multiple-choice item was written into three versions that varied in the amount of multiple-response options presented. Analyses of these items' characteristics were designed to ascertain whether the Combinational-Format Items were more difficult or discriminating than their regular multiple-choice item counterparts. Experiment II addressed the question of the amount of cognitive processing needed to respond to items containing single- and multiple-response options.

Experiment I

Method

Subjects and setting. Three intact undergraduate Foundations of Education classes at a large midwestern university, composing a total of 120 students, participated as part of a class presentation on measurement issues. All subjects were tested in their regular classroom by one of the experimenters.

Materials. Two sets of ten items each were identified from previous American College Testing Programs Social Studies experimental tryout units. Set One had correct answers that were single-option and Set Two had correct answers that were all inclusive ("All of the above"). Within each set, each item was rephrased into three versions that were labeled A, B, and C. Version A items were written in the usual linear form whereby the first three of the four alternatives presented a separate, independent statement for the examinee to consider and the last alternative was "All of the above." One of these alternatives was keyed as the correct answer: for Set One, the keyed answer was foil a, b, or c, while for Set Two, the keyed answer routinely was d, "All of the above." For Version B, each item was rephrased into a I, II, III multiple response item format such that the alternatives were in an one-to-one parallel with their Version A counterparts. Thus, the alternatives for Version B (single-response) were "I only," "II only," "III only," in the a-, b-, and c-alternatives positions, and "I, II, and III" in the d-alternative position. For Set One items, the correct answer was either "I only," "II only," or "III only" while Set Two items always had "I, II, and III" as the keyed answer. Therefore, the rephrasing of items from Version A to B allowed for the investigation of a strict translation from a linear-type presentation into a multiple-response format. In Version C (multiple-response), each item's alternatives were again rephrased,

but this time some combinational alternatives were presented, although they were never keyed as the correct answer. The conversion of items from Version B to C, then, allowed for the presentation of conceptually complex combinations of options in the alternative set. A total of sixty items were developed (20 * 3 versions). The items were randomized within the test booklets in such a way that each student had one of the three versions of each item, and one form of each item was represented in each test booklet.

An example of each of the three versions of an item is presented below.

Version A

Which of the following would accurately distinguish between the parliamentary democracies of Western Europe and the United States system of government?

a. In parliamentary democracies, a majority vote in the legislative branch is necessary for enacting legislation.

b. In parliamentary democracies, members of the legislative branch are elected by the citizenry.

* c. In parliamentary democracies, it is impossible for one party to control the legislative branch while another controls the executive.

d. All of the above

Versions B and C

Item Stem and Options:

Which of the following would accurately distinguish between the parliamentary democracies of Western Europe and the United States systems of government?

I. In parliamentary democracies, a majority vote in the legislative branch is necessary for enacting legislation.

II. In parliamentary democracies, members of the legislative branch are elected by the citizenry.

III. In parliamentary democracies, it is impossible for one party to control the legislative branch while another controls the executive.

Version B alternatives	Version C alternatives
a. I only	* a. III only
b. II only	b. I and III only
* c. III only	c. II and III only
d. I, II, and III	d. I, II, and III

Procedure. The two forms of the test were randomly assigned to the students in each class. Students were informed that the test was a sample of social studies items of the type developed for use in a standardized test. The students were asked to read each item carefully and to choose the best alternative for each item. After all students had completed their tests, the materials

were collected and the experimenter continued a discussion of item-answering strategies for multiple-choice tests.

Results

Item characteristics (difficulty and discrimination indices) were calculated for the items within the experimental unit in which they appeared. The analysis of the effects of item format concentrated on the characteristics of difficulty and discrimination. The items were then sorted into the appropriate research form (A, B, or C) for subsequent analyses. The data were subjected to a two factor (Answer Status: single or all-inclusive; Form: A, B, C) repeated measures analysis of variance. Form was deemed a within factor since the basic item stem and options remained constant across Forms A, B, and C, while Answer Status was a between factor. The analyses were done separately for item difficulty (proportion of students selecting the keyed answer) and item discrimination (point biserial correlation coefficients). For item difficulty, no significant Form effect was found, F (2,38) = 0.47, $p > .50$. Mean item difficulty for Form A was 0.49 with a standard deviation of 0.19. Mean item difficulty for Form B was 0.47 with a standard deviation of 0.17 while Form C had a mean and standard deviation for item difficulty of 0.46 and 0.17, respectively. Likewise, nonsignificance was found for the analysis of item discrimination, F (2,38) = 0.47, $p > .50$. Item discrimination means and standard deviations for the three forms are as follows: Form A: M = 0.34,

SD = 0.17; Form B: M = 0.33, SD = 0.16; Form C: M = 0.33, SD = 0.19.

Discussion

Given the lack of significance found in this study for the phrasing of items in straight linear, single-response, or multiple-response formats, for either single-answer of all-inclusive-answer items, there appears to be little reason to prefer one item phrasing over the other for these items. The items showed similar item difficulty and discrimination values regardless of their response-option status.

In order to determine whether the Combinational-Format Item was somehow different in the processing demands it places on the examinee, a second study was performed. Britton and his colleagues (e.g., Britton, Glynn, Meyer, and Penland, 1982; Britton, Holdredge, Curry & Westbrook, 1979; Britton, Pina, Davis, & Westhausen, 1978) have developed a methodology to assess the amount of cognitive capacity persons bring to bear during various kinds of information-processing tasks. The procedure developed by Britton and his colleagues involves determining students' reaction times to auditory stimuli (in Britton's terms, the monitoring for auditory tones is referred to as the "secondary task") while they are engaged in various other, primary tasks (e.g., reading an essay). In a recent study (1982), for example, Britton and Tesser determined that as demands on students' knowledge about a processing task

increased, so too did students' reaction times to the auditory stimuli. And, while other hypotheses have been suggested to account for this phenomenon (see Britton et al., 1982), Britton and his colleagues have proposed that the increased reaction times indicate that greater amounts of subjects' cognitive capacities are being devoted to the primary tasks. When greater amounts of capacity are employed in a primary task, less capacity is available for the secondary task (monitoring for tones) and so reaction time increases.

It appears that Britton's methodology, originally used to determine cognitive capacity usage among students reading prose materials, could be adapted to examine relative demands for cognitive processing for answering various types of test items.

Experiment II

Procedures and Methodology

For this study, Versions B and C from Experiment I were used. That is, only the versions that presented the alternatives in a I, II, III format were selected. Version B items presented options in a single-combination format: "I only," "II only," "III only," and "I, II, III." Version C provided some alternatives that combined the options, such as "I and II only." Therefore, subjects were asked to consider double combinations of options in Version C. Note, however, that double options were never keyed as the correct response to an item. The items were randomized within two test forms in such a way that each student was presented one of the two

versions of each item and one version of each item was presented in each form.

Subjects and setting. Forty-three undergraduates enrolled in educational psychology courses at a large midwestern university volunteered to participate for course credit. All data were gathered in individual sessions during three weeks in a small conference room set aside for that purpose.

Apparatus. Presentation of the materials was via a TRS-80 Model III microcomputer. Tones were presented via an attached Atcher mini-amplifier speaker.

Procedure. Each subject was seated in front of the microcomputer and told that he/she would be shown a series of multiple-choice questions dealing with general knowledge of social studies. Each subject was further informed that the answers to the questions were to be marked on an answer sheet furnished for that purpose. Then subjects were told that during their reading they would occasionally hear a tone (at this point the experimenter had the machine emit a tone through the speaker) and that when they did, they were to press the space bar as rapidly as possible. After the instructions, subjects were shown a sample question. The stem of each question (including the three options) was called up by pressing a key. Then, after the item was read, the alternatives were added to the screen by pressing the key again. When subjects had answered a question, they could remove the question they had been working on

and move to the next stem by pressing a key. Subjects were given a practice item and three practice tones. The experimenter then answered any questions and made sure that subjects were positioned with a thumb on the space bar.

Six tones were presented in each subject's test. The tones occurred only when the alternatives were called up at from 1.5 second to 3.0 second intervals after the alternatives were presented. Three of the tones were randomly distributed in the alternatives for Version B items and three tones were randomly distributed in the alternatives for the Version C items. Each subject received a new random ordering of the tones. The reaction times to the tones were recorded internally by the TRS-80.

Results

The reaction time data for each version of the items (B and C) were averaged for each subject, and then a related samples t test was performed. The results, t (42) = 4.13, $p < .001$, indicated significantly longer reaction times for Version C items (M = 0.49, SD = 0.26) than for Version B items (M = 0.29, SD = 0.94).

Because the computer presentation in Experiment II may have affected the validity of the item characteristics computed in Experiment I for these items, item characteristics (difficulty and discrimination) were again analyzed as they had been in Experiment II. For item difficulty, no significant Version effect was found, F (1,19) = 2.65, $p > .10$. Also, no significant effect for Version

was found when item discrimination indices were analyzed, F (1,19) = 0.53, $p > .50$. Means and standard deviations for item characteristics by Version are as follows: Version B, difficulty index, M = 0.44, SD = 0.10; discrimination index, M = 0.35, SD = 0.14; Version C, difficulty index, M = 0.52, SD = 0.13; discrimination index, M = 0.37, SD = 0.13.

Discussion

The results of the reaction time analysis in Experiment II indicated that significantly longer reaction times were required for Version C items than for Version B items. From the perspective of Britton and Tesser (1982), this result indicates that more of the subjects' cognitive capacities were required for Version C items than for Version B items. This increase in cognitive processing could involve more reading by the examinees or more use of an "executor processor" (i.e., the brain) in juggling and keeping track of the combinations, or even be indicative of higher-level processing while working on the item. The results of the analyses of item characteristics were consistent with the results of Experiment I; that is, no significant differences were observed between item types on measures of difficulty or discrimination. Therefore, conflicting results were found between the reaction time and item characteristics analyses; item characteristics of difficulty and discrimination appear not to be affected by the combinational presentation of the options, whereas, combinational options resulted

in more processing demands than did noncombinational ones. Thus, it would seem that measures of cognitive processing among students may provide a more sensitive measure of the multiple dimensions of multiple-choice test items than the more typical analyses of item characteristics.

General Discussion

Although no differences in item difficulty or discrimination indices were found for the differential display methods of options for the items, items that require combinational reasoning (like those presented in Version C) were shown to require a greater amount of reaction time to a secondary task. This suggests that, if the arguments of cognitive psychology researchers are used, the Combinational-Format Items do require more cognitive demands for processing than do single-response items. The methodology from cognitive psychology does appear to have promise for assessing cognitive processing demands of various item types.

With regard to the questions this study set out to address, it seems that item characteristics of difficulty and discrimination are not affected by the use of multiple-response items. However, based on the reaction time to a secondary task, Combinational-Format Items do appear to be different in the processing demands placed on the subject.

References

Albanese, M. A. (1982). Multiple-choice items with combination of correct responses: A further look at the Type K format. Evaluation and the Health Professions, 5, 212-228.

Albanese, M. A., Kent, T. H., Whitney, D. R. (1979). Cluing in multiple-choice test items with combinations of correct responses. Journal of Medical Education, 54, 948-950.

Britton, B. K., Glynn, S. M., Meyer, B. J. F., & Penland, M. S. (1982). Effects of text structure on use of cognitive capacity during reading. Journal of Educational Psychology, 74, 51-61.

Britton, B. K., Holdredge, T. S., Curry, C., & Westbrook, R. D. (1979). Use of cognitive capacity in reading identical texts with different amounts of discourse level meaning. Journal of Experimental Psychology: Human Learning and Memory, 5, 262-270.

Britton, B. K., Piha, A., Davis, J., & Westhausen, E. (1978). Reading and cognitive capacity usage: Adjunct question effects. Memory and Cognition, 6, 266-273.

Britton, B. K. & Tesser, A. (1982). Effects of prior knowledge on use of cognitive capacity in three complex cognitive tasks. Journal of Verbal Learning and Verbal Behavior, 21, 421-436.

Harasym, P. H., Norris, D. A., & Lorscheider, F. L. (1980). Evaluating student multiple-choice responses. Evaluation and the Health Professions, 3, 63-84.

Kolstad, R. K., Wagner, M. J., Kolstat, R. A., & Miller, E. G. (1983). The failure of distractors on complex multiple-choice items to prevent guessing. Educational Research Quarterly, 8, 44-50.